RESEARCH COMMUNICATION

A novel human airway mucin cDNA encodes a protein with unique tandem-repeat organization

Viswanathan SHANKAR,* Michael S. GILMORE,† Ronald C. ELKINS† and Goverdhan P. SACHDEV*‡

*College of Pharmacy and †College of Medicine, University of Oklahoma Health Sciences Center, P.O. Box 26901, Oklahoma City, OK 73190, U.S.A.

Highly specific affinity-purified polyclonal antibodies against deglycosylated human tracheobronchial mucin was used to select immunoreactive clones from a Uni-ZAP cDNA expression library prepared from normal human tracheal mRNA. The largest of three positive clones, designated pAM1, which reacted strongly with the polyclonal antibodies, was further characterized. Sequence analyses revealed a partial 941 bp cDNA that encoded a 313-amino-acid polypeptide. Bases 3–892 consisted of imperfect 41-nucleotide tandem repeats (CCAGGAGGGGAC-ACCGGGTTCACGAGCTGCCCACGCCCTCT) that encoded a unique polypeptide with two types of consensus repeats, TSCPRPLQEGTRV and TSCPRPLQEGTPGSRAAHALSR-RGHRVHELPTSSPGGDTGF. The overall composition of the deduced amino acid sequence matched that expected for a mucin protein core and is rich in serine, threonine, proline, glycine and alanine (≈ 51 %). Northern blots probed with the mucin cDNA exhibited intense polydisperse hybridization bands with RNA isolated from normal human trachea and cystic-fibrosis bronchus. The data indicate that mucin encoded by clone pAM1 represents a unique type of peptide organization which has not been described in mucin cDNAs reported thus far.

INTRODUCTION

Mammalian respiratory, gastrointestinal and reproductive tracts are protected by mucus secretions, the major components of which are the mucin glycoproteins. Mucins constitute a heterogeneous group of high-molecular-mass, polydisperse, highly glycosylated macromolecules secreted by the goblet cells of the epithelia and the submucosal-gland cells. Mucins are overproduced in chronic obstructive pulmonary disease such as chronic bronchitis, asthma and cystic fibrosis (CF), presumably due to hyperplasia of mucus-secreting cells (mucus-cell metaplasia) and/or overexpression of specific mucin gene(s) [1].

Our understanding of the structure of mucins has advanced rapidly in the past few years with the isolation and characterization of cDNA clones encoding mucin polypeptide cores. To date, at least seven human mucin genes (*MUC1–MUC7*) have been identified [2–8]. All the mucins appear to share common characteristics, including extended arrays of tandemly repeated sequences within the protein core, which is rich in potential O-glycosylation sites and has a high content of serine, threonine, proline, glycine and alanine (> 50 % of the total amino acids). The tandem amino acid repeat units vary in length from as few as eight amino acids per repeat unit in *MUC5* [6] to 169 amino acids in *MUC6* [7]. The tandem repeat domain(s) are flanked on either side by non-repeat regions.

Recently Dufosse et al. [9] reported a new type of organization in mucin genes. Four overlapping mucin clones isolated from a tracheobronchial cDNA library showed degenerate 87 bp tandem repeats that enigmatically encoded a polypeptide devoid of repeat sequences because of numerous additions and deletions in the nucleotide sequence. Here we report yet another novel type of airway-mucin gene organization. We have isolated a partial cDNA from a normal human tracheal library that encompasses almost perfect 41 bp tandem repeats and encodes a polypeptide with two types of consensus-peptide tandem repeats. This mucin type is expressed at high level in human airways.

EXPERIMENTAL

Purification of human tracheobronchial mucin, deglycosylation and production of antiserum

A detailed protocol for the purification of the major human tracheobronchial mucin component (HTM-1) was described previously [10]. Further purification of mucin component HTM-1 was carried out by CsCl-density-gradient centrifugation, essentially as described previously for minor mucin component HTM-2 [11]. The mucin was found to be highly pure and devoid of any low-molecular-mass protein contaminants. Purified mucin was deglycosylated by exposure to anhydrous HF for 3 h at room temperature, conditions which have been shown to remove essentially all of the oligosaccharide side chains [12].

Deglycosylated mucin (100 μg in 0.5 ml of complete Freund's adjuvant) was injected subcutaneously into New Zealand White rabbits on day 1. Two additional booster injections of 50 μg of deglycosylated mucin in incomplete Freund's adjuvant were administered on days 23 and 35 [13]. Antisera were collected, and the titre was measured against deglycosylated mucin using a standard e.l.i.s.a. [14]. The IgG fraction of the polyclonal antiserum was purified on a Protein A–Superose column (FPLC; Pharmacia). Antigen-specific antibodies were isolated from the IgG fraction by immunoaffinity purification on a column prepared by coupling deglycosylated mucin to CNBr-activated Sepharose CL4B using standard methods [15].

Ouchterlony double diffusion, dot-blot assays and e.l.i.s.a established the specificity of antiserum to deglycosylated HTM-1. In these experiments, the antiserum reacted strongly with deglycosylated HTM-1 and only weakly with native mucin. In

Abbreviations used: HTM-1, major human tracheobronchial mucin component; HTM-2, minor component; CF, cystic fibrosis.

‡ To whom correspondence should be addressed.

The sequence data will appear in the GenBank/EMBL/DDBJ Databases under the accession no. U04799.

e.l.i.s.a., the antiserum reacted with deglycosylated HTM-1 at dilutions up to 10^{-4}. In competitive e.l.i.s.a., the deglycosylated HTM-1 at a level of 80 ng gave 50% inhibition. Human serum albumin, hyaluronic acid, chondroitin sulphate and keratan sulphate did not inhibit the binding of antibody to deglycosylated HTM-1, even at a level of 10 μg. Thus the high titre of the antiserum, absence of reactivity against other glycoconjugates and the ability of deglycosylated HTM-1 to inhibit antibody binding in competitive e.l.i.s.a. confirmed the specificity of the antiserum.

Construction and screening of the cDNA library

A normal human trachea was obtained from an autopsy within 30 min *post mortem*, frozen immediately in liquid N_2 and used for mRNA isolation employing the PolyATract mRNA isolation kit (Promega). The integrity of the mRNA was checked by Northern-blot hybridization using as a control human-β-actin-specific cDNA. The cDNA library was constructed in the Uni-ZAP vector (Stratagene) using an oligo(dT) linker-primer to initiate first-strand synthesis as recommended by the manufacturer. The primary library was found to contain 0.9×10^6 primary recombinants with an average insert size of 1.8 kb and was subjected to one round of amplification before screening.

The cDNA library was screened using the affinity-purified polyclonal antibodies to deglycosylated human tracheal mucin (HTM-1) using standard protocols [16]. Three positive plaques that reacted strongly with the antiserum were obtained after screening about 0.25×10^6 plaques and were purified to clonality by further rounds of screening. The pBluescript vector containing the inset was obtained by *in vivo* excision using helper phage R408.

Restriction mapping and sequencing

The nucleotide sequence of the 941 bp insert of pAM1 was determined by the dideoxy-chain-termination method using α-^{35}S-dATP and Sequenase 2.0 (U.S. Biochemicals) on both single- and double-stranded templates. Sequence information was obtained from both strands of the DNA in their entirety. When necessary, dITP was substituted for dGTP to resolve band compressions and ambiguities. Nested deletions were obtained using exonuclease III and S1 nuclease (Erase-A-Base System; Promega). Sequences not obtained from nested deletions were obtained by subcloning restriction-endonuclease-generated fragments into pBluescript.

Northern-blot analyses

Total RNA was isolated from normal human trachea and CF bronchus using the RNAgents kit from Promega. Human brain, small intestine and liver RNA was obtained from Clontech. RNA samples were subjected to electrophoresis in a 1%-agarose/formaldehyde denaturing gel, transferred to GeneScreen Plus (NEN/du Pont) nylon membrane and probed with cDNA probes under high-stringency conditions as suggested by the manufacturer. cDNA probes were labelled to high specific radioactivity with α-[^{32}P]dCTP using a random-primer labelling kit (DECAprime II; Ambion).

Sequence homology searches

Nucleotide and protein sequence homology searches with sequences in the GenBank database were conducted using the University of Oklahoma Genetic Computer Group programs.

RESULTS

Initial screening of the cDNA library yielded three antibody-positive plaques which were purified to clonality by further rounds of screening. Partial sequence analyses of these three clones from both ends of the insert DNA revealed a tandemly repeating nucleotide sequence from either end in two of the shorter clones, while the largest of them (designated pAM1) had an identical tandem repeat sequence at the 5′-end with additional non-repeat sequences at the 3′-end. This clone was further characterized and included a single open reading frame encoding 313 amino acids. Bases 3–892 contained degenerate 41 bp tandem repeats (Figure 1) that encoded two types of consensus repeat peptides (Figure 2). The peptide sequence TSCPRPLQEGTPG-SRAAHALSRRGHRVHELPTSSPGGDTGF was encoded by three of the 41 bp repeats in tandem, while deletions of two bases in the otherwise virtually perfect 41 bp repeats led to 39 nucleotide repeats encoding the peptide TSCPRPLQEGTRV. The deletion in position 809 of repeat unit 20 (Figure 1) caused a shift in reading frame, leading to disruption of the peptide repeat (Figure 2, amino acids 270–297). Recently, Dufosse et al. [9] have isolated human airway-mucin cDNA clones where deletions or additions of nucleotides caused numerous shifts in the reading frame, thus completely disrupting the repetitive peptide structure. To the best of our knowledge, the peptide organization represented by clone pAM1 has not been observed in any mucin cDNAs known thus far. The overall composition of the deduced amino acid sequence matched that expected for a mucin protein core. The combined content of serine, threonine, proline, glycine and alanine was $\approx$ 51% of the total amino acids. Interestingly, cysteine made up $\approx$ 4% of the total amino acids, with one cysteine residue per repeat unit.

Northern-blot analyses with the 941 bp cDNA insert revealed intense polydisperse hybridization bands typical of mucin probes extending from greater than 9 kb to around 500 bases (Figure 3). The mucin cDNA hybridized to RNA from normal human

```
CCACGAGCTGCCCACGTCCTCTCCAGGAGGGGACACCGGGTTC   43
  ACGAGCTGCCCACGCCCTCTCCAGGAGGGGACACCGGGTTC   84
  ACGAGCTGCCCACGTCCTCTCCAGGAGGGGACACCGGGTTC  125
  ATGAGCTGCCCACGCCCTTTCCAGGAAGGGACCCCGGGTTC  166
  ACGAGCTGCCCACGTCCTCTCCAGGAAGGGACCCCGGGTTC  207
  ACGAGCTGCCCACGTCCTCTCCAGGAAGGGACCCCGGGTTC  248
  ACGAGCTGCCCACGTCCTCTCCAGGAAGGGACCC-GGGT-C  287
  ACGAACTGCCCACGTCCTCTCCAGGAAGGGACCCCGGGTTC  328
  ACGAGCTGCCCACGTCCTCTCCAGGAGGGGACACCGGGTTC  369
  ACGAGCTGCCCACGCCCTCTCCAGGAAGGGACCCCGGGTTC  410
  ATGAGCTGCCCACGTCCTCTCCAGGAAGGGACCC-GGGT-C  449
  ACGAACTGCCCACGCCCTCTCCAGGAGGGGACCC-GGGT-C  488
  ACGAGCTGCCCACGTCGTCTCCAGGAAGGGACCC-GGGT-C  527
  ACGAGCTGCCCACGTCCTCTCCAGGAAGGGACCC-GGGT-C  566
  ACGAACTGCCCACGCGCTCTCCAGGAGGGGACACCGGGTTC  607
  ACGAGCTGCCCACGCCCTCTCCAGGAAGGGACCCCGGGTTC  648
  ACGAGCTGCCCACGTCCTCTCCAGGAGGGGACACCGGGTTC  689
  ACGAGCTGCCCACGTCCTCTCCAGGAGGGGACACCGGGTTC  730
  ACGAGCTGCCCACGCCCTCTCCAGGAGGGGACACCGGGTTC  771
  ACGAGCTGCCCACGTCCTCTCCAGGAAGGGACCC-GGGT-C  810
  ACGAGCTGCCCACGTCCTCTCCAGGAGGGGACACCGGGTTC  851
  ACGAGCTGCCCACGCACTTTCCAGGAAGGGACCCCGGGTTC  892
aggtctcctgccggcccacatcgtgcctttgtgtaaatcagaagaaaga  941
```

Figure 1 Nucleotide sequence of clone pAM1

The nucleotide sequence is arranged in such a way that the tandem repeats are aligned. Gaps in the otherwise perfect 41 bp repeats are denoted by '—'. The non-repeat region at the 3′ end of the clone is denoted in lower-case letters.

```
TSCPRPLQEGTPGSRAAHALSRRGHRVHELPTSSPGGDTGF   41
MSCPRPFQEGTPGSRAAHVLSRKGPRVHELPTSSPGRDPGF   82
TSCPRPLQEGTRV                               95
TNCPRPLQEGTPGSRAAHVLSRRGHRVHELPTPSPGRDPGF  136
MSCPRPLQEGTRV                              149
TNCPRPLQEGTRV                              162
TSCPRRLQEGTRV                              175
TSCPRPLQEGTRV                              188
TNCPRALQEGTPGSRAAHALSRKGPRVHELPTSSPGGDTGF  229
TSCPRPLQEGTPGSRAAHALSRRGHRVHELPTSSPGRDPG   269
                          HELPTSSPGGDTGF   283
TSCPRTFQEGTPGS                             297

GLLPAHIVPLCKSEER                           313
```

Figure 2 Deduced amino acid sequence of clone pAM1

The two types of peptide repeats encoded by pAM1 are aligned under each other in the order in which they are encoded. The underlined sequence represents that encoded by the non-repeat portion.

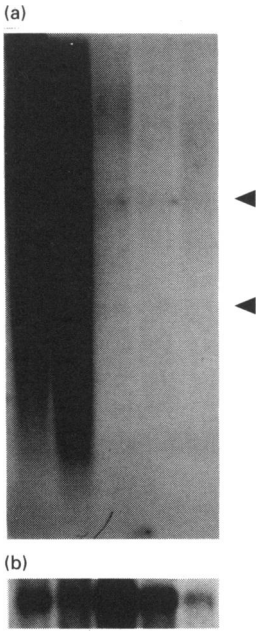

Figure 3 Northern-blot analyses of total RNAs (10 g/lane)

Lanes 1–5 represent RNA from CF bronchus, normal human trachea, brain, small intestine and liver respectively. (**a**) Shows the hybridization pattern observed with the 941 bp cDNA insert of pAM1 as a probe. Arrowheads indicate the position of the 28 S and 18 S ribosomal RNA bands. (**b**) Shows the same blot reprobed with a cDNA for human glyceraldehyde-3-phosphate dehydrogenase to check RNA integrity.

trachea and from CF bronchus, but not to RNA from human brain, small intestine and liver. Furthermore, 41- or 39-base oligonucleotides identical with the tandem repeat sequences, or a 48-base oligonucleotide corresponding to bases 894–941 of the non-repeat portion of pAM1, gave hybridization signals similar to that obtained with the 941 bp cDNA. This provided strong evidence for these regions being an integral part of the same message.

Computer searches for similarities of the nucleotide and deduced amino acid sequence showed no significant homology with mucins MUC1–MUC7 or to any other proteins in the database.

DISCUSSION

Cloning and sequencing data from mucin (MUC1–MUC7) studies thus far has resulted in the emergence of a number of consensus features characteristic of mucins. First, mucin protein cores contain a high percentage (> 50%) of serine, threonine, proline, glycine and alanine. In that encoded by pAM1 the combined content of these five amino acids is ≈ 51%. Secondly, the protein backbone contains a number of tandemly repeating amino acids encoded by tandemly repeated nucleotide sequences. The only exception among human mucin genes is the recent report by Dufosse et al. [9], that numerous deletions and additions occurred in an otherwise perfect 87 bp tandem repeat sequence resulting in a protein backbone with no repeats. This observation indicated a new type of organization in mucin genes. Clone pAM1 appears to represent another novel type of organization within a mucin genes, wherein degenerate 41 bp tandem repeats encode a protein that exhibits two types of tandem-repeat sequence. When pAM1 was used to probe Northern blots, a polydisperse hybridization band was observed. Further, the expression of this mucin may be tissue-specific, since only human tracheal/bronchial RNA showed positive hybridization signals, while brain, small intestine and liver were negative. However, it remains to be established whether other mucin-secreting tissues express this novel airway mucin. Taken together, these data and the fact that pAM1 was isolated by using highly specific polyclonal antibodies to human tracheal mucin, provide compelling evidence that the cDNA represented in pAM1 encodes a novel mucin polypeptide.

The nucleotide and deduced amino acid sequence of pAM1 showed no sequence similarity to that of any other human or animal mucin, and no significant similarity was observed with any other gene sequence in the GenBank database. This is not surprising, since previously described mucin sequences also lack sequence similarity at the nucleotide or protein level in their tandem repeats. While regions outside of the tandem repeat domain, especially at the C-terminus, have shown some degree of conservation [8], the limited sequence information on the 3′ end of pAM1 showed no significant matches. While the mechanisms that lead to the deduced amino acid sequence of clone pAM1 are reminiscent of those described for *MUC5B* [9], it is premature to state whether polypeptides encoded by pAM1 and *MUC5B* represent different regions of a single gene product. We are currently in the process of determining the chromosomal localization of the gene encoding pAM1, which may shed more light on this aspect.

On the basis of sequence and physicochemical properties, secretory mucin monomers ware viewed as containing heavily glycosylated and rigid central repetitive domains flanked by unique sequences [17]. The unique sequences are known to contain numerous cysteine residues that apparently link mucin monomers into oligomers via disulphide bonds. While cysteine residues have not been observed in previously described mucins within the repeat regions, pAM1 seems unique in that it encodes a peptide with one cysteine residue per repeat unit. It is premature to speculate whether these cysteine residues play any role in mucin oligomer formation via disulphide linkages. It would be interesting to elucidate the structure of the full-length mucin corresponding to pAM1 and see if the regions flanking the repeat

domain show high content of cysteine residues as observed in mucin MUC2 [18].

Interestingly, except in the case of *MUC7* [8], all human mucin cDNA clones initially isolated from either oligo(dT)- or random-primed cDNA libraries lacked a complete 3′ end, including a poly(A) tail [2–7]. The isolation of partial clones lacking a poly(A) tail in the present study is thus consistent with the results reported previously. Although no direct evidence has been presented to date, it appears that this may be related to the stability of the regions flanking the tandem-repeat domains during propagation in a prokaryotic host [18,19].

In summary, we have isolated a partial cDNA that encodes a unique human airway-mucin glycoprotein. Characterization of this novel cDNA has revealed a new type of mucin peptide organization. Both normal and CF tracheobronchial tissue showed high level of expression of this novel mucin, which may be airway-specific. Thus the protein encoded by clone pAM1 appears to be a major mucin component secreted in the human airways. Using the 941 bp cDNA insert of pAM1 as a probe, we are in the process of obtaining longer/full-length clones. Further characterization of these clones may shed more insights on the novel mucin glycoprotein. Also it will be of considerable interest to determine whether the expression of this novel mucin type is altered in the airways of individuals with CF and other obstructive pulmonary diseases.

This work was supported in part by a grant from National Heart, Lung and Blood Institute (HL34012).

REFERENCES

1 Basbaum, C. B. and Finkbeiner, W. E. (1989) in Lung Cell Biology (Massaro, D., ed.), pp. 37–80, Marcel Dekker, New York

2 Gendler, S. J., Taylor-Papadimitriou, J., Duhig, T., Rothbard, J. and Burchell, J. (1988) J. Biol. Chem. **263**, 1280200–12823

3 Gum, J. R., Byrd, J. C., Hicks, J. W., Toribara, N. W., Lamport, D. T. A. and Kim, Y. S. (1989) J. Biol. Chem. **264**, 6480–6487

4 Gum, J. R., Hicks, J. W., Swallow, D. M., Lagace, R. L., Byrd, J. C., Lamport, D. T. A., Siddiki, B. and Kim, Y. S. (1990) Biochem. Biophys. Res. Commun. **171**, 407–415

5 Porchet, N., Van Cong, N., Dufosse, J., Audie, J. O., Guyonnet-Duperat, V., Gross, M. S., Denis, C., Degand, P., Bernheim, A. and Aubert, J. P. (1991) Biochem. Biophys. Res. Commun. **175**, 414–422

6 Crepin, M., Porchet, N., Aubert, J. P. and Degand, P. (1990) Biorheology **27**, 471–484

7 Toribara, N. W., Roberton, A. M., Ho, S. B., Kuo, W.-L., Gum, E., Hicks, J. W., Gum, J. R., Byrd, J. C., Siddiki, B. and Kim, Y. S. (1993) J. Biol. Chem. **268**, 5879–5885

8 Bobek, L. A., Tsai, H., Biesbrock, A. R. and Levine, M. J. (1993) J. Biol. Chem. **268**, 20563–30569

9 Dufosse, J., Porchet, N., Audie, J. P., Guyonnet-Duperat, V., Laine, A., Van-Seuningen, I., Marrakchi, S., Degand, P. and Aubert, J. P. (1993) Biochem. J. **293**, 329–337

10 Shankar, V., Naziruddin, B., Reyes de la Rocha, S. and Sachdev, G. P. (1990) Biochemistry **29**, 5856–5864

11 Padhye, N. V., Naziruddin, B., Reyes de la Rocha, S. and Sachdev, G. P. (1991) Biochim. Biophys. Acta **1077**, 332–338

12 Byrd, J. C., Lamport, D. T. A., Siddiki, B., Kuan, S.-F., Erickson, R., Itzkowitz, S. H. and Kim, Y. S. (1989) Biochem. J. **261**, 617–625

13 Vaitukaitis, J. L. (1981) Methods Enzymol. **73**, 46–52

14 Desai, V. C., Naziruddin, B., Graves, D. C., Reyes de la Rocha, S. and Sachdev, G. P. (1991) Hybridoma **10**, 285–296

15 Harlow, E. and Lane, D. (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

16 Sambrook, J., Fritsch, E. F. and Mainiatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

17 Gum, J. R. (1992) Am. J. Respir. Cell Mol. Biol. **7**, 557–564

18 Gum, J. R., Hicks, J. W., Toribara, N. W., Rothe, E. M., Lagace, R. E. and Kim, Y. S. (1992) J. Biol. Chem. **267**, 21375–21383

19 Gendler, S. J., Lancaster, C. A., Taylor-Papadimitriou, J., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Lalani, E. and Wilson, D. (1990) J. Biol. Chem. **265**, 15286–15293

Received 24 February 1994/17 March 1994; accepted 23 March 1994